Regularization for Deep Learning

Sargur N. Srihari
srihari@cedar.buffalo.edu

What is Regularization?

- Central problem of ML is to design algorithms that will perform well not just on training data but on new inputs as well
- Regularization is:
  - "any modification we make to a learning algorithm to reduce its generalization error but not its training error"
  - Reduce test error even at the expense of increasing training error

Some Goals of Regularization

- Many forms of regularization available
  - Major efforts are to develop better regularization
- Put extra constraints on objective function
  - They are equivalent to a soft constraint on parameter values
    - Result in improved performance on test set
- Some goals of regularization
  1. Encode prior knowledge
  2. Express preference for simpler model
  3. Needed to make underdetermined problem determined

Regularizing Estimators

- In Deep Learning, regularization means regularizing estimators
- Involves increased bias for reduced variance
  - Good regularizes reduces variance significantly while not overly increasing bias

Model Types and Regularization

- Three types of model families
  1. Excludes the true data generating process
     - Implies underfitting and inducing high bias
  2. Matches the true data generating process
  3. Overfits
     - Includes true data generating process but also many other processes
- Goal of regularization is to take model from third regime to second

Importance of Regularization

- Overly complex family does not necessarily include target function, true data generating process, or even an approximation
- Most deep learning applications are where true data generating process is outside family
  - Complex domains of images, audio sequences and text true generation process may involve entire universe
    - Fitting square hole (data generating process) to round hole (model family)

What is the Best Model?

- Best fitting model obtained not by finding the right number of parameters
- Instead, best fitting model is a large model that has been regularized appropriately
- We review several strategies for how to create such a large, deep regularized model

Regularization Strategies

1. Parameter Norm Penalties
   - (L^2- and L^1- regularization)
2. Norm Penalties as Constrained Optimization
3. Regularization and Under-constrained Problems
4. Data Set Augmentation
5. Noise Robustness

More Regularization Strategies

6. Semi-supervised learning
7. Multi-task learning
8. Early Stopping
9. Parameter tying and parameter sharing
10. Sparse representations
11. Bagging and other ensemble methods
12. Dropout
13. Adversarial training
14. Tangent methods